Explanation*

MICHAEL REDHEAD

1. Introduction

In what sense do the sciences explain? Or do they merely describe what is going on without answering why-questions at all. But cannot description at an appropriate 'level' provide all that we can reasonably ask of an explanation? Well, what do we mean by explanation anyway? What, if anything, gets left out when we provide a so-called scientific explanation? Are there limits of explanation in general, and scientific explanation, in particular? What are the criteria for a *good* explanation? Is it possible to satisfy all the desiderata simultaneously? If not, which should we regard as paramount? What is the connection between explanation and prediction? What exactly is it that statistical explanations explain? These are some of the questions that have generated a very extensive literature in the philosophy of science. In attempting to answer them, definite views will have to be taken on related matters, such as physical laws, causality, reduction, and questions of evidence and confirmation, of theory and observation, realism versus anti-realism, and the objectivity and rationality of science. I will state my own views on these matters, in the course of this essay. To argue for everything in detail and to do justice to all the alternative views, would fill a book, perhaps several books. I want to lead up fairly quickly to modern physics, and review the explanatory situation there in rather more detail.

2. Why-questions: the D-N model as providing a necessary condition for explanation

'Why' is ambiguous.[1] If someone asserts that the earth is round, and I ask 'Why', I may be asking for a reason why you believe that the earth is round. What is your evidence for the assertion? Typically you may refer to *consequences* of the earth being round, horizon phenomena, the shape of the shadow of the earth on the face of the moon during an

* I acknowledge the warm hospitality of the Wharfedale farmhouse where this paper was written.

[1] Cf. Hospers (1967), 240.

eclipse, and so on. The evidence you cite supports or confirms the hypothesis that the earth is round, justifies you in believing that the earth is round. But I may also be asking for an explanation: why *is* the earth round? Is it just a brute fact or can you *explain* it, in a way which confers understanding and removes my perplexity in the face of just reciting the brute fact? Typically you will cite the roundess of the earth as itself a *consequence* of other propositions in a manner which contributes to my understanding. This is the famous deductive model of explanation.[2] Is deducibility either sufficient or necessary for a scientific explanation? Well, it is obviously not sufficient. Most trivially X is deducible from X, and we can hardly allow X to explain itself. But I am going to claim that it is necessary. This is to fly in the face of most of the post-Hempelian literature on explanation.[3] There are two sorts of counter-examples to necessity commonly cited. First, volcanoes—we cannot deduce when they will erupt, but, so it is said, we can explain an eruption *post hoc* in terms of known geological process. But that is to confuse an explanation-sketch with a full-blooded scientific explanation.[4] If we knew enough about the distribution of stress in the rocks and the laws governing mechanical rupture we could make the prediction. If we do not possess all the relevant information then we are not in a position to give a scientific explanation, that is full or complete. That is not to say that partial explanations may not confer a measure of understanding, but they do not measure up to the scientific ideal involving strict deducibility, i.e., in the volcano example, predictability.

There is an important ambiguity here we must be clear about. Many physical systems are governed by deterministic laws in the sense that *exact* specification of initial and boundary conditions fixes the later physical state uniquely, but the prediction is unstable in the sense that any error, *however small*, introduces divergent behaviour in specifying the future state of the system. So, in practice, and even in principle, *we* cannot compute the prediction. But from a 'God's eye' point of view everything is fixed. In this paper we are not concerning ourselves primarily with the pragmatics of explanation, and hence when we talk of predictability, we mean this in the ontological 'God's eye' sense.

But some events cannot be predicted. Does this mean they cannot be explained? I think the appropriate answer to this question is to bite the

[2] The classic statement of the D-N model of explanation is Hempel and Oppenheim (1948)—reprinted in Hempel (1965).

[3] Influential critiques of the Hempelian approach include Scriven (1962) and Salmon (1971). A comprehensive survey of the literature on explanation is provided by Achinstein (1983).

[4] Cf. Hempel (1965), 416.

bullet and say 'yes'. So what do probabilisitic or statistical explanations achieve? Well, they enable us to deduce and hence to explain the limiting relative frequencies with which events of a given kind turn up in a long-run repetition of the set-up producing the phenomenon. But they cannot explain what happens on a particular occasion. It is as simple as that!

In particular they cannot explain in the strict deductivist sense the relative proportions in a *finite* sample. Suffice it to say that these proportions can be used to provide a rational estimation of what can be explained, i.e. the probabilities, in accordance with the usual statistical procedures.

There is an enormous literature on statistical explanation, arguing whether the probability of the explanandum has to be high in the presence of the explanans[5] or merely higher than it would be in the absence of the explanans.[6] In the strict *scientific* sense, I regard these discussions as irrelevant. Again I am not denying that there are senses of promoting understanding, or removing perplexity, other than that produced by the scientific ideal. But it is with the scientific ideal that we are concerned.

But if deducibility is admitted to be necessary for scientific explanation, how are we to fill out the conditions to achieve sufficiency. That is a much more difficult task, to which I now turn.

3. The Circularity Objection

In the deductive-nomological (D-N) model of Hempel the explanans cites one or more scientific laws. In the usual schematic fashion adopted by philosophers of science, let us represent a typical scientific law in the universally quantified form $\forall x(Px \rightarrow Qx)$—succinctly all Ps are Qs.

If *a* is a P, i.e. Pa is true, then we seek to explain why *a* is a Q by deducing Qa from the premisses

$$\forall x(Px \rightarrow Qx) \tag{1}$$

$$Pa \tag{2}$$

Thus: from (1) by Universal Instantiation

$$Pa \rightarrow Qa \tag{3}$$

Whence, from (2) and (3) by *modus ponens*

$$Qa$$

Why is this thought to be explanatory? In (1) the implication as we have written it is material implication. On a Humean (regularity) view of

[5] See Hempel (1965), 376ff.

[6] Cf. Salmon (1971).

laws that is all there is to (1).[7] It is true in virtue of all its instances being true. But if (1) depends for its truth on the truth of (3), and this, given the premiss Pa, must turn on the truth Qa. So is not the argument completely circular? The truth of Qa, given Pa, is grounded in the truth of a universal stratement, whose truth is grounded in the truth of Qa, the very fact we are trying to explain. What this amounts to is that (1) is nothing more nor less, on the Humean account, than a compendium of all the instances (3). (In the case of a finite variety of instances the universal law is indeed nothing else than the conjunction of its instances.) On the Humean account the instances are 'loose' (there is no cement!) so effectively the Hempelian model, under this interpretation of law, amounts to the assertion that facts only explain themselves.

But this whole argument hinges on the assumption that the universal law is only supported or confirmed by the totality of evidence which would make the law (deductively) true! If we think that (1) may be supported in some *inductive* sense by instances other than (3) then the circularity would be satisfactorily mitigated. But here we are backing ourselves straight into the problem of induction. That induction is not *deductively* valid as a mode of inference, committing as it does the fallacy of affirming the consequent, must of course be admitted. So in our example, if we are to avoid circularity, we must acknowledge that the explanans is never known definitely to be true, and this is an obvious, but unavoidable, defect in scientific explanations. For a Popperian, of course, the matter is much worse than that. Finite evidence never supports to any degree the conjectural truth of universal laws quantified over infinite domains. So a Popperian expects, in so far as he allows himself any expectations, that an essential part of the explanans, in any scientific explanation, is definitely false (although not currently known to be false!). Does this render scientific explanations irrational?—a conclusion much advertised by the critics of Popperism. The answer depends very much on what we count as rational. Is it rational to aim at the impossible? Can we not rationally *accept* scientific laws, without believing them to be true, provided they have been subject to the severest available *criticism*?[8]

Let us turn from these deep methodological issues to consider whether a necessitarian account of natural laws helps with the question of explanation. On the question of evidence not at all. After all, that was

[7] The Humean view of laws is most recently defended in Swartz (1985). More sophisticated Humeans of the Ramsey-Lewis stripe incorporates a 'more for less' criterion for natural laws that links closely with our later discussion of Friedman's views on explanation and unification. A good exposition of the Ramsey-Lewis approach is provided in Armstrong (1983), chapter 5.

[8] For a sustained defence of neo-Popperian rationality see Watkins (1984).

Hume's original point. We have no epistemological access to the idea of necessary connection. But let us take an ontological standpoint. If laws of nature involve nomological necessity, however analysed, over and above the merely factual material implication, would this not account for how Pa being the case is the ground for Qa being the case? The ground for Qa being grounded in Pa is the necessary connection between P'ness and Q'ness on one popular understanding of these matters.[9] Of course, we can press on to query the ground for *that* ground. We shall return to the question of ultimate or self-supporting explanations presently. But there are other matters we want to deal with.

Firstly, superfluous aspects of explanation: we deduced Qa from (1) and (2), but we could also deduce it from (1), (2) and

God's in his Heaven (4)

We must check that (4) is idle, that we could equally deduce Qa from (1), (2) and the negation of (4).

Then there is the question of explanations which could be correct but as a matter of fact are not. Example: Fred who is shot through the head, after he has died as a result of being stabbed through the heart! The shot could have been the correct explanation of his death, but in deciding whether it actually is, we need to attend to all the relevant circumstances. The law linking shooting through the head with death is more accurately rendered as linking the shot with a transition from life to death. As Fred is already dead, this more fully amplified version of the law is no longer applicable and cannot be cited in explaining Fred's death.[10] Again the scientific ideal assumes that all the relevant circumstances are being cited.

So far we have given some indication of what constitutes an explanation in science, but what constitutes a *good* explanation, when is one explanation better than another?

4. Good Explanations: Unification

There is first of all the element of suprise, of unexpectedness, the 'Aha' factor.[11] That is no doubt related to the criterion of non-circularity, that the explanandum in no way presents itself as what we take to be the

[9] Cf. Armstrong (1983).

[10] For a critical discussion of whether *ceteris paribus* clauses and provisos can in principle or in practice be spelled out in the required detail see Grünbaum and Salmon (1988).

[11] See Feigl (1949).

evidence for the explanans. In practice good explanations, by this dimension of appraisal, arise at the intersection of several universal laws, all of which are necessary to deduce the explanandum.[12] Iron sinks in water, not just because all solids with density greater than that of water sink in water, or even with density greater than that of liquid in which it is placed, but better, because the difference in density is associated via Archimedes Principle with an imbalance of weight versus buoyancy, which shows, from the laws of mechanics, the direction in which the iron will move. The sinking of the iron is not just a special case of a more inclusive law, and so would not be cited as direct evidence for any of the laws stated separately.

But there is another important aspect that this example brings out. What we described as the better explanation involves a series of laws, which can be used to deduce and hence explain many other facts in the field of hydrostatics or with appropriate extension of the theoretical apparatus, the enormous richness of hydrodynamic phenomena from waterfalls to the breaking of waves on the seashore, from whirlpools to the lift and drag of an aerofoil. This points to the very important unification aspect of explanation. The world at the surface level of immediate experience appears very complicated, very rich in diverse phenomena with no apparent connection. But at a 'deeper' theoretical level can all this diversity get reduced to a few interlocking explanatory principles? This has always provided an ideal of theoretical progress in science, the ideal of unification. There is no doubt that the history of modern physics has provided examples of increasing unification in our fundamental theories. But it is important to be clear as to what is being claimed. There are a number of distinct senses of unification that need to be distinguished.[13]

Firstly, there is the question of the inter-relatedness, or 'working together' of the explanatory nexus. Supose we have two sorts of phenomena, P_1 and P_2, which stand for the sets of lawlike regularity governing the phenomena (at the immediate 'empirical' level of everyday physicalist discourse) and suppose that P_1 and P_2 are explained by theories T_1 and T_2. Then $P_1 \cup P_2$ is certainly explained by $T_1 \wedge T_2$ since any member p_1 of P_1 can be deduced from T_1 and any member p_2 of P_2 can be deduced from T_2. In a trivial sense there are new predictions that can be deduced from $T_1 \wedge T_2$ but not from either theory separately, viz. conjunctions like $p_1 \wedge p_2$. But there are no *interesting* novel predictions that arise from $T_1 \wedge T_2$. In one sense of unification, a unified explanation of $P_1 \cup P_2$ would arise from a theory T in which there was no partition of the axioms which separately yielded all the (interesting)

[12] See Nagel (1961), 34ff.
[13] Cf. Redhead (1984).

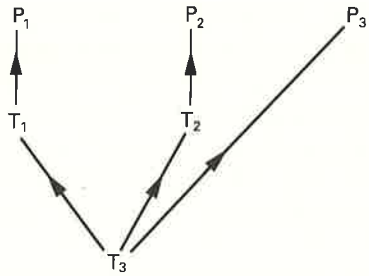

Figure 1

predictions. There would be predictions that required the interlocking working together of the axioms. Typically a unified explanation T of $P_1 \cup P_2$ would also predict new phenomena arising from the interactive effect of the axioms comprising T. Such an explanation would not only be unified but could be in a sense 'deeper' than T_1 and T_2. Suppose we denote the new phenomena by P_3, then we can illustrate the situation we have in mind very schematically as shown in Figure 1. T_3 not only explains T_1 and T_2 that originally accounted for P_1 and P_2, but makes new prediction P_3 and this is done in such a way that T_3 provides a unified account of $P_1 \cup P_2 \cup P_3$. There are many complications associated with working out this idea of unity-cum-depth. These are admirably treated in Watkins' 1984-monograph *Science and Scepticism*. Firstly T_3 may correct T_1 and T_2, not just unify them, and then the idea of increased empirical content becomes formally problematic. Watkins deals with this by his method of counterparts in which incompatible statements are 'matched up' according to appropriate rules. Then there is the problem that a theory such as T_3 which may be unified under one axiomatization may become non-unified under an alternative axiomatization. To deal with this problem Watkins resorts to a notion of 'natural axiomatization', satisfying rules that prevent unnecessary proliferation of axioms and defeat proposed unification-defeating reaxiomatizations. In addition to independence and non-redundancy requirements, Watkins has a rule demanding segregation of axioms containing only theoretical terms, but more importantly there is a non-molecularity requirement, stating that an axiom is impermissible if it contains a proper component which is a theorem of the axiom set, or becomes one when its variables are bound by the quantifiers that bind them in the axiom. Finally there is a decomposition requirement specifying that if the axiom set can be replaced by an equivalent one that is more numerous, without violating the other rules, it should be.

I want to use Watkins' idea of a natural axiomatization to solve a vexing problem that now arises. Why does T_3 remove our perplexity about $P_1 \cup P_2 \cup P_3$? Part of the answer lies in showing that these

apparently unconnected phenomena are in fact related via the unified derivation from T_3 (under a natural axiomatization). But we also want to say that T_3 is in some intuitive sense simpler than $T_1 \wedge T_2$ together with P_3 itself. An obvious approach here is just to count the number of axioms $N_A(T_3)$ in a natural axiomatization of T_3 and check whether this is less than $N_A(T_1)+N_A(T_2)+N_A(P_3)$. (Note that Watkins assumes the underlying logic and mathematics is already 'given', so we are concerned with a *finite* number of non-logical and non-mathematical axioms. These may be regarded as part of the specification of a class of models in the Sneed-Stegmüller structuralist approach to theories, although Watkins himself seems to subscribe to the standard 'statement' view of theories'.) What do we mean by $N_A(P_3)$? Well it is the number of axioms in a natural axiomatization of the phenomena comprised in P_3. This in effect means counting the number of laws in P_3. But we must be careful that the laws are expressed in a manner allowed by Watkins' rules. For example, as Friedman (1974) noted, any statement Q can be expressed as the conjunction of N statements, for any N. Thus write

$$
\begin{aligned}
Q &\equiv P_1 \wedge (P_1 \rightarrow Q) \\
&\equiv P_2 \wedge (P_2 \rightarrow P_1) \wedge (P_1 \rightarrow Q) \\
\hline
&\equiv P_{n-1} \wedge (P_{n-1} \rightarrow P_{n-2}) \wedge (P_{n-2} \rightarrow P_{n-3}) \ldots \wedge (P_1 \rightarrow Q)
\end{aligned}
$$

Where $P_1, P_2 \ldots P_{n-1}$ is a descending chain of increasingly weak logical consequences of the statement Q. Such rewritings for law statements Q would be eliminated immediately by Watkins' rules against inessential proliferation of axioms.

We can also compare $N_A(T_1)$ with $N_A(P_1)$ and $N_A(T_2)$ with $N_A(P_2)$ to see how a reduction in the number of laws left unexplained may already have been achieved at the level of introducing T_1 and T_2, and compare with the further reduction effected by T_3.

On this account the reduction in our perplexity in the face of P_1, P_2, P_3 corresponds just to the reduction to $N_A(T_3)$ of those laws of nature which we have to accept without explanation. This was the intention behind Friedman's (1974) approach to explanation, but the account he gave was technically quite incoherent, as shown by Kitcher (1976). Kitcher himself rejected the simple idea of counting laws in an explanatory framework in favour of (effectively) counting what he calls patterns of explanation (see Kitcher (1981) for details). I believe myself that the original Friedman approach is much more straightforward and perspicuous, if it can be rescued from the Kitcherian strictures by employing the Watkins natural axiomatization approach in the way I have described.

There is one important matter we must attend to. We have spoken of laws of nature comprised in the axioms of T_3 explaining laws of nature comprised in the phenomena P_1, P_2 and P_3. But we must be clear as to what we are going to allow as a law of nature. It expresses a regularity, that all items of a certain kind, the subject S of the law, subject to certain conditions C, behave according to some property P. The problem in counting laws is concerned with the question of how general S needs to be and how specific C. If we allow for the maximum generality in S and the minimum specificity in C we could easily rule out electromagnetic laws, for example, other than: all charged bodies obey Maxwell's equations! Achinstein in his (1971) argues that we may refer to a universal generalization as a law if at some period in the history of science it is not known (or believed) that generalization is possible, to a more inclusive law. But this approach introduces what many philosophers regard as an unacceptable historical relativism in place of an objective criterion. Perhaps we should just accept that all generalizations that can be deduced as theorems in an axiomatic-deductive framework should count as laws. The question at issue is also related to the vexed notion of natural kinds, that the extension of the subject-predicate in a law should not be the result of some arbitrary and conventionally imposed system of classification, that the subject-predicate should correspond to a 'genuine' universal. At all events, while there may be borderline cases, such as the much-discussed example of Kepler's laws, where the decision in these matters is somewhat controversial, I believe that sufficient liberality in identifying generalization as laws will permit the notion of unification to be a useful and important one in discussions of the scientific explanation of laws. The essential point is that unification delimits what we must accept without explanation.

In many discussions of explanation the point is made that the explanans must comprise some distinguished set of explanatory principles. On some accounts[14] the explanans must have an analogy with phenomena with which we are familiar in everyday experience, that perplexity in the face of the explanandum is removed by exploiting our familiarity with the analogy. Crudely we may require the explanans to be in some sense picturable. This requirement, while clearly exemplified in examples such as the nineteenth century mechanical aether models, is not characteristic of explanations in modern theoretical physics. While pictures, such as the Feynman diagrams in elementary particle physics, are an aid to keeping track of complicated computational procedures, they are actually potentially very mislead-

[14] See Campbell (1920) and Hesse (1966).

ing in affording what we may term physical understanding of what is going on.[15]

On other accounts the privileged set of explanatory principles is relativized to what in a particular theoretical-conceptual scheme are regarded as 'natural', as not requiring explanation. We shall return to this requirement in a moment in discussing the relevance of causality to questions of explanation. But for the moment, we would merely note that the superiority of the unification approach of Friedman is that an arguably objective criterion can be presented that characterizes progress in the explanatory endeavour in science.

One of the most favoured approaches to unification in modern physics has been through the process of micro-reduction. This raises a number of issues which we now turn to.

5. Micro-reduction

The micro-reduction programme sees the microscopic world as composed of unobservable 'atomic' units whose properties serve to explain in a unified way the whole range of macroscopic phenomena. Unification now incorporates the *additional* requirement that the explanatory principles only cite properties of the microscopic entities involved in the explanans.

The question of the terminus of explanation now presents itself as the alternative between ultimate atoms and an infinite regress of 'Chinese boxes', in which micro-entities are themselves resolved into micro-micro constituents *ad infinitum.* Molecules are resolved into atoms, atoms into electrons and atomic nuclei, nuclei into protons and neutrons, nucleons into quarks . . . One must not think that micro-reduction necessarily eliminates the autonomous branches of macroscopic physics, in favour of a physics of elementary particles. The identity statements characteristic of scientific reductions may best be understood as contingent, but law-like coextensionalities. One can imagine worlds where hot bodies exist, but temperature is not a measure of mean energy of the constituent molecules. That may not be our world, but it is a possible world, whose very conceivability shows that the identity statements in reduction schemes are not analytic in character.[16]

While it is true that the major trend in modern theoretical physics has been micro-reductive, there are some caveats that must be entered.

(1) On the orthodox Copenhagen interpretation of quantum mechanics, the macroscopic world of classically described apparatus and

[15] See Redhead (1988a) for further discussion of this point.

[16] The view expressed here should be contrasted with that canvassed by Causey (1977).

experimental set-ups must be presupposed in analysing the properties of those very micro-entities which make up or constitute the macroscopic objects. In a sense the reduction instead of descending linearly towards the elementary particles, moves in a circle, linking the reductive basis back to the higher levels.[17]

(2) Even on non-orthodox interpretations of quantum mechanics of the hidden-variable variety, there is a choice between non-local action-at-a-distance between elementary particles and a holistic conception of non-separability, that in so-called entangled states it does not make sense to attribute properties to separate 'particular' entities, independently of a prior understanding of the properties of the whole composite system.[18]

(3) In resolving the problem of measurement in quantum mechanics some have held that human consciousness plays an essential role in resolving the ambiguity in 'pointer readings' predicted by straightforward application of the quantum-mechanical formalism. This would again make the reductive hierarchy Psychology → Biology → Chemistry → Physics → Elementary Particles bite at its own tail. We shall return to the question of human consciousness in scientific explanations shortly in discussing so-called anthropic explanations.

(4) In the bootstrap approach to elementary particle physics that was very popular in the 1960s, a democratic approach to composite entities, constituent entities and 'force-carrying' entities was proposed. Every entity could potentially play any of these three roles, so that the analysis of composite entities into a few 'aristocratic' constituents in the typically micro-reductive fashion was rejected.[19]

Despite these reservations, micro-reduction still seems a viable option, if taken as in point (2) with unorthodox interpretations that allow some sort of action-at-a-distance between micro-entities. This poses potential problems in reconciling quantum mechanics with special relativity. But perhaps only if the action-at-a-distance is of a causal character. This is the first point at which we have mentioned the notion of cause, and we want now to say something about the view that the only 'genuine' explanations of events occurring in the physical world are causal explanations.

6. Causal Explanations

What are causes? There is no consensus among philosophers, although they generally regard them as an important matter in any metaphysical

[17] For further discussion see Redhead (1987a).
[18] Cp. Redhead (1987b).
[19] For details see Redhead (1980a).

comprehension of the world. The surprising thing is that physicists long ago gave up the notion of cause as being of any particular interest![20] In physics the explanatory laws are laws of functional dependence, how one physical magnitude is related in a regular (and law-like on the necessitarian account) fashion with another physical magnitude. The pressure and volume of an ideal gas are related at constant temperature by Boyle's law. The pressure does not cause the volume, or the volume the pressure, they just coexist in a manner regulated by Boyle's law. But of course there are dynamical laws showing how the state of physical system depends on the time. Consider an object falling from rest with a constant acceleration g. After time t the distance s it has fallen is given by Galileo's law, $s=\frac{1}{2}gt^2$, so at t_1 the distance is $s_1=\frac{1}{2}gt_1^2$ and at a later time t_2 the distance is $s_2=\frac{1}{2}gt_2^2$. In what sense does the event consisting in the object having fallen a distance s_1 at t_1 cause the object to reach s_2 at t_2?

All we have is the relation

Position at t_1 + Galileo's law $\Rightarrow$ Position at t_2

The position at t_1 can hardly be cited as the cause of the position at t_2. Surely the cause must be something that is introduced to account for why the position of the object has changed. Well, is not the answer to cite the acceleration? But the acceleration is just *defined* by the kinematic relationship expressed in Galileo's law. So can the cause be Galileo's law? But Galileo's law is not another event which happens at t_1, 'causing' the particle to move with constant acceleration and so to reach s_2 at t_2. It is just an expression of that acceleration. So should we not retreat to citing the force, such as gravity, which 'causes' the acceleration. But the idea that forces 'cause' bodies to move is a very anthropomorphic notion. What we actually have in physics is a force law, such as the inverse square law of gravitational attraction, which relates, via Newton's second law, the acceleration of the body to the relative location of other bodies such as the earth.

Instead of $s=\frac{1}{2}gt^2$, we have, in idealized approximation, $s=\frac{1}{2}(GM/R^2)\,.\,t^2$, where M is the mass of the earth, R its radius and G is a new gravitational constant. So we are back with a regularity connecting s with t, but also now with M and R. But the force of gravity has been eliminated between the force law and Newton's second law. Many metaphysicians want to resist this elimination. There really is a force which then causes the motion. But the sense in which they mean this is that the force causes the body to change its *natural state of motion*, as

[20] For the historical background to the changing attitude of physicists to the notion of cause see Kuhn (1971), reprinted in Kuhn (1977). Compare also the classic paper of Russell (1917).

specified by Newton's first law, that if a body is not moving, its natural state is one of perpetual rest. There is a definite sense in which the accelerated motion is not 'natural' and hence has to be caused. But modern physicists would reject this distinction. Acceleration under gravity is just as 'natural' a change, as perpetual rest or uniform motion in the circumstances appropriate to those descriptions. All of this is vividly illustrated in general relativity where all motion is geodesic motion (in four-dimensional space-time). If space-time is not flat this geodesic motion is a close approximation to the old Newtonian motion with acceleration induced by a force. But there is no such thing as a non-natural motion. To most physicists the old-fashioned idea of cause arises from the idea of our interfering in the natural course of events, pushing and pulling objects to make them move and so on. In modern physics there are just regularities of one sort or another. They are all 'natural' and hence leave no room for causation. It is true that some natural regularities can be analysed in the manner of one event impinging on a contiguous one, so transmitting action in a continuous train of contiguous action. This is true of many field theories, which play such a prominent role in modern physics. But this causal method of articulating what is going on in a field theory should not be taken seriously by philosophers. It is, at best, a crutch to the understanding of non-physicists!

But what, you may say, of causality principles employed by physicists? That effects cannot precede their causes and so on. What physicists mean here is that law-like regularities in nature must not be such that if we, *as free agents*, interfered in the natural course of events, we could alter past events, in such a way as to induce inconsistencies in the specification of present states of affairs. But backwards causation and forward causation into the past, via closed time-like world lines in general relativistic space-times, are not *a priori* prohibited, only in so far as they could lead to causal loop paradoxes. Theories which exhibit these 'anomalous' features express constraints on present states of affairs, that must make them consistent with 'causal loop' restrictions. There is less contingency in the world than we had suspected, there are events like shooting our own grandmothers that we are not free to bring about, on pain of rendering our physical theories logically inconsistent.

7. Quantum-mechanical Correlations and Causation

In the preceding section I have put forward the view that I believe most physicists would have concerning causation. This is reinforced by the fact that causal processes in the old classical sense are not contemplated

in the orthodox interpretations of our most fundamental theory, quantum mechanics. But what about the unorthodox hidden-variable interpretations? Since these are inspired by classical intuitions, many philosophers feel that questions of causality need to be discussed and evaluated. Perhaps the world really is ultimately to be understood, to be explained, in the language of causal networks.[21] I want to discuss this possibility briefly in the face of EPR correlations and the Bell arguments.

How should one explain correlations, understood in the sense of the probabilistic dependence of one event on another? On a causal view of explanation, there are two options. Firstly, a direct causal link, and secondly, a common cause which screens off one event from the other. Of course these options are not mutually exclusive—a common cause may be overlaid by a direct cause. Now, in the case of the correlations contemplated in the Bell version of the Einstein–Podolsky–Rosen experiment in its Bohm spin-correlation version, a common cause explanation of the correlations leads to the constraint known as the Bell inequality which is violated by the predictions of quantum mechanics (and also by experiment). So the common cause must be overlaid by a direct causal interaction, but since the events comprised in spin-components in specified directions taking values on the two 'wings' of the experiment may be at space-like separation, this poses problems for special relativity as mentioned at the end of section 5.

A way out of this impasse has been proposed by the present author (1988b) in terms of an explicit explication of Shimony's idea of passion-at-a-distance.[22] What has been shown is that the Bell-type correlations can be understood in terms of a probabilistic connection between the two events, which lacks the 'robustness' necessary for a causal connection. By robustness one means here that either event screens off the other from the way it was produced. The Bell correlations are not robust in this sense. The situation is rather like a stick which maintains a correlation, namely the relative distance of the two hands of someone holding on to the two ends of the stick, but which is such that if one end of the stick is 'jerked' by someone else, it mysteriously adjusts to a new length! Of course robustness may be restored by citing who jerked the end of the stick. If Jack jerks the stick it goes to one length, if John jerks it to another. But suppose the length depends also on who knocked Jack, with the result that he jerked the stick, and who knocked Bill who knocked Jack *ad infinitum* . . . This would be a marshmallow world where no adequate notion of causation as an explanation of the correlation between the ends of the stick could be rescued. The Bell correla-

[21] Cf. Salmon (1984).
[22] Shimony (1984).

tions are not like that; robustness is lost 'at a stroke' so to speak. But is talk of passion-at-a-distance in this situation genuinely explanatory? Is it not just giving a name to the brute fact of the correlations? I think not, because we are pointing here to a precise characterization that is exemplified in the Bell situation, and hence enables us to get a grip on what sort of situation we are dealing with. In the last section we saw how physicists have recourse to laws of functional dependence, but the philosopher may well be interested in classifying these laws into the coexistence of properties of natural kinds, dynamical laws, causal laws, 'passion' laws, and so on. Additional understanding of the physical situation may result from philosophical analysis of the category of law statements involved in the explanans.

8. Anthropic Explanation

In the recent literature, there has been considerable discussion of the explanatory virtues of the so-called anthropic principle in understanding the values of fundamental physical constants, such as the mass-ratios of elementary particles, or the dimensionless couplings characterizing the strengths of interactions.[23]

The basic idea is very simple. It is a remarkable fact that if the values of such fundamental constants differed significantly from the values they actually have, then the necessary conditions for the emergence of carbon-based life would not obtain, and we could not now exist, to speculate on the explanation of why these constants have the values they do! Or, arguing contrapositively, the fact of our own conscious existence implies within a very narrow 'window' the values of the fundamental constants. It is now argued that the fact of our existence is the explanation of why these constants have the values they do, modulo, of course, the fundamental laws governing the behaviour of micro-entities, other than those specifying the values of the constants in question.

Let us immediately distinguish a weak and strong version of this anthropic principle. The weak version says just that: our existence allows us to infer the values of certain fundamental constants. Our existence is, if you like, an *indicator* of what values these constants have. But the strong version of the anthropic principle claims that our existence not only allows us to infer the values of the constants, but is the explanatory ground for the values they have. We have already discussed briefly the relationship between explanation and prediction

[23] See Barrow and Tipler (1986) for a very extensive discussion of the physics involved.

(or more generally indication if we do not want to stress the temporal aspect). We have claimed that all complete scientific explanations *were* indicators. But what about the converse? Does mere indication qualify as explanation? The answer to this question is generally agreed to be no. Counter-examples commonly cited include the length of the shadow of a flagpole that indicates for us, but does not explain, the length of the flagpole; the period of a pendulum which indicates the length of the string, but does not explain why the string has that length; the barometer that indicates, but does not explain, the weather, and so on. There has also been much discussion as to the exact reasons that underlie the immediate intuition that the examples cited are not explanatory. The unification approach to explanation can help us here. Science looks for unified patterns of explanation.[24] In the case of the flagpole and the string of the pendulum, we explain their length by appeal to the way they were produced, i.e., on a micro-reductive account, citing the molecular structure of the component materials and how this allowed division by a saw or pair of scissors or whatever, to the dimension in question. And this pattern of explanation is available for all questions of the form 'Why is such-and-such the size it is?' Similarly for the barometer, explanation of the weather and all other hydrodynamic phenomena follow the unified pattern of deriving consequences from the macroscopic laws of hydrodynamics together with appropriate initial conditions, or at a deeper level, from the kinetic theory of fluids. Now, in discussing the implication afforded by micro-reductions we have already stressed that the explanatory apparatus comrises the micro-entities and their properties, thus coralling off a few deep fundamental principles that, at any rate for the time being, are not susceptible of explanation.

But in the anthropic mode of explanation we substitute the orthodox and unified explanatory scheme with another non-integrated scheme, the fundamental laws at the deep theoretical level plus something up at the macroscopic level, viz. our own existence. And notice furthermore that the existence of dragonflies or crocodiles would do just as well as our own existence to provide this sort of explanation. Indeed generalizing the pattern of anthropic explanation could lead us to cite any fact as the ground for any other fact that can be deduced from it, thus trivializing the notion of explanation in the same way that the simple-minded D-N model allows. In addition, on the basis of counting the number of constituents in the explanatory scheme as a measure of unification, we can argue that the existence of a living organism unpacks into a very large number of facts, the instantiation of all those law-governed biochemical processes necessary to life, as compared with

[24] Cf. Kitcher (1981).

the relatively small number of constants whose values are being explained.

The fact of the matter is that the strong anthropic principle arises from an initial prejudice in favour of reconstituting Man at the centre of the Universe, rather than examining the whole matter dispassionately in the light of those desiderata we have detailed for a satisfactory scientific explanation.

Notice carefully I am not saying that the orthodox scientific account, that the fundamental laws *plus* the constants of nature are the ontological ground for the *possibility* of our existence, is true in a way which has been established beyond any shadow of doubt, and that the anthropic account, that our existence modulo the laws of nature is the ontological ground for the constants having the values they do, is false. I am arguing only what it is rational, scientifically speaking, to accept. Scientific rationality in no way depends on establishing indubitable truths about the world. In the context of the truth-status of scientific laws, rather than questions of what is the ontological ground for what, as we penetrate below the surface level of appearance the whole account becomes essentially conjectural, but these are not conjectures made in some intellectual vacuum, they are controlled and made criticizable by the possible disagreement of their empirical consequences with experiment. This is what makes the acceptance of these conjectures scientifically rational, not their demonstrated and certifiable truth. It is rather like the creationist versus evolutionist controversy. The objection to creationism is not that science has shown it to be false, but that it is not consonant with scientific rationality to accept it. In the explanation case the situation is admittedly a little different, since even the negative control of experiment is not available to bear on the issue. The decisive point is that the orthodox unified account does confer a greater measure of understanding in the sense we have explained, whereas the anthropic account could only be regarded as superior on an assumption of teleological necessity, to which we have no empirical access. Notice that we are not attempting ourselves to argue in favour of causal necessity as opposed to teleological necessity. As we intimated in section 6, it is philosophers who are led to such considerations, perhaps even on the basis of believing, incorrectly, that physics has something to contribute in this area! We maintain that any such arguments must be essentially metaphysical in character.

9. Conclusion

It is time to sum up and answer a few remaining questions. The ideal of scientific explanation is a matter of logical deduction, given a unified set

of deep explanatory principles that are themselves accepted, for the time being, without explanation. But of course the ideal of scientific explanation is one for ongoing improvement. Perhaps from the fundamental laws of microphysics, by some consistency criterion, it will turn out that the constants of nature are tightly constrained or even uniquely determined. But even then we would still have the task of explaining the laws themselves at a still more fundamental level. At some stage scientific explanations always turns into description—'That's how it is folks'—there is no *ultimate* terminus in science for the awkward child who persists in asking why! I do not believe the aim of some self-vindicating *à priori* foundation for science is a credible one.

Then there is the question of delimiting contingency in the world. On a necessitarian view of laws, contingency is usually understood in terms of the contrasted non-necessity of initial and boundary conditions. But in matters of cosmology that sort of distinction may not be the right one to draw. The no-boundary universe of Hawking and Hartle is an example, where essentially there are only laws![25]

Another matter deserving of mention is the role of symmetry principles, which are really metalaws constraining the form of laws themselves.[26] Many explanations in modern physics bypass the detailed laws by invoking directly the metalaws. But their status again is not self-vindicating. The discovery of the violation of mirror symmetry in the physics of weak interactions was one of the crowning triumphs of modern science.

Some people have argued that the orthodox ideal we have been expounding and defending should be given up, and replaced by an account which lies closer to scientific practice.[27] Now, it is quite true that the fundamental laws of microphysics are quite useless as a practical basis for deducing, and hence explaining many phenomena. The fundamental theories are mathematically much too intractable. So, in practice, physicists use all kinds of approximate and indeed inconsistent 'models', to discuss the properties of complicated systems such as atomic nuclei or lasers.[28] But I have been concerned with the scientific ideal. The fact that, for practical reasons, it cannot be implemented, does not detract from its status as an ideal, something we should try for, rather than substituting the messy 'real' physics as something which, instead of falling short of the ideal standard, should itself be elevated into the standard for scientific explanation.

[25] See Hawking (1988) for a popular exposition.
[26] See Redhead (1975).
[27] Cf. Cartwright (1983).
[28] See Redhead (1980b).

In his book *The View from Nowhere* (1986), Tom Nagel has argued that science in eschewing subjectivity, does not tell the whole story. To give a proper account of human actions and intentions we need the subjective view, the view from somewhere. But in its proper place and sphere the traditional methods of objective science have proved extraordinarily successful—there is no reason in my view to think that modern micro-reductive physics throws inevitable doubt on those methods and their motivating ideal of the progressive unification of science.

References

Achinstein, P. 1971. *Law and Explanation*, (Oxford: Clarendon Press).

Achinstein, P. 1983. *The Nature of Explanation*, (Oxford University Press).

Armstrong, D. M. 1983. *What is a Law of Nature*, (Cambridge University Press).

Barrow, J. D. and Tipler, F. J. 1986. *The Anthropic Cosmological Principle* (Oxford: Clarendon Press).

Campbell, N. R. 1920. *Physics: The Elements*, (Cambridge University Press).

Cartwright, N. 1983. *How the Laws of Physics Lie*, (Oxford: Clarendon Press).

Causey, R. 1977. *Unity of Science*, (Dordrecht: Reidel).

Feigl, H. 1949. 'Some Remarks on the Meaning of Scientific Explanation', in H. Feigl and W. Sellars (eds.) *Readings in Philosophical Analysis*, 510–14.

Friedman, M. 1974. 'Explanation and Scientific Understanding', *Journal of Philosophy*, **71**: 5–19.

Grünbaum, A. and Salmon, W. C. 1988. *The Limitations of Deductivism* (Berkeley: University of California Press).

Hawking, S. W. 1988. *A Brief History of Time*, (London: Bantam Press).

Hempel, C. G. 1965. *Aspects of Scientific Explanation*, (New York: The Free Press).

Hempel, C. G. and Oppenheim, P. 1948. 'Studies in the Logic of Explanation', *Philosophy of Science*, **15**: 135–75.

Hesse, M. B. 1966. *Models and Analogies in Science*, (Notre Dame University Press).

Hospers, J. 1967. *An Introduction to Philosophical Analysis*, 2nd ed. (London: Routledge & Kegan Paul).

Kitcher, P. 1976. 'Explanation, Conjunction and Unification', *Journal of Philosophy*, **73**: 207–12.

Kitcher, P. 1981. 'Explanatory Unification', *Philosophy of Science*, **48**: 507–31.

Kuhn, T. 1971. 'Les Notions de causalité dans le developpement de la physique', *Etudes d'Épistémologie Génétique*, **25**: 7–18.

Kuhn, T. 1977. *The Essential Tension*, (University of Chicago Press).

Nagel, E. 1961. *The Structure of Science*, (London: Routledge & Kegan Paul).

Nagel, T. 1986. *The View from Nowhere*, (Oxford University Press).
Redhead, M. L. G. 1975. 'Symmetry in Intertheory Relations', *Synthese*, **32:** 77–112.
Redhead, M. L. G. 1980a. 'Some Philosophical Aspects of Particle Physics', *Studies in History and Philosophy of Science*, **11**: 279–304.
Redhead, M. L. G. 1980b. 'Models in Physics', *British Journal for the Philosophy of Science*, **31**: 145–63.
Redhead, M. L. G. 1984. 'Unification in Science', *British Journal for the Philosophy of Science*, **35**: 274–9.
Redhead, M. L . G. 1987a. 'Whither Complementarity?', in N. Rescher (ed.) *Scientific Inquiry in Philosophical Perspective*, (Lanham: University Press of America), 169–82.
Redhead, M. L. G. 1987b. *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*, (Oxford: Clarendon Press).
Redhead, M. L. G. 1988a. 'A Philosopher Looks at Quantum Field Theory', in H. R. Brown and R. Harré (eds.), *Philosophical Foundations of Quantum Field Theory* (Oxford: Clarendon Press), 9–23.
Redhead, M. L. G. 1988b. 'Nonfactorizability, Stochastic Causality, and Passion-at-a-Distance', to appear in J. Cushing and E. McMullin (eds.), *Philosophical Consequences of Quantum Theory.*
Russell, B. 1917. 'On the Notion of Cause', reprinted in *Mysticism and Logic*, (London: Allen and Unwin), 180–208.
Salmon, W. 1971. *Statistical Explanation and Statistical Relevance* (Pittsburg University Press).
Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*, (Princeton University Press).
Scriven, M. 1962. 'Explanations, Predictions and Laws', in H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science III*, 170–230.
Shimony, A. 1984. 'Controllable and Uncontrollable Non-Locality', in S. Kamefuchi *et al*. (eds.) *Proceedings of the International Symposium: Foundations of Quantum Mechanics in the Light of New Technology*, (Tokyo: Physical Society of Japan), 225–30.
Swartz, N. 1985. *The Concept of Physical Law*, (Cambridge University Press).
Watkins, J. W. N. 1984. *Science and Scepticism* (Princeton University Press).